Defense Business Systems Common Vocabulary - Common Core

Dennis E. Wisnosky

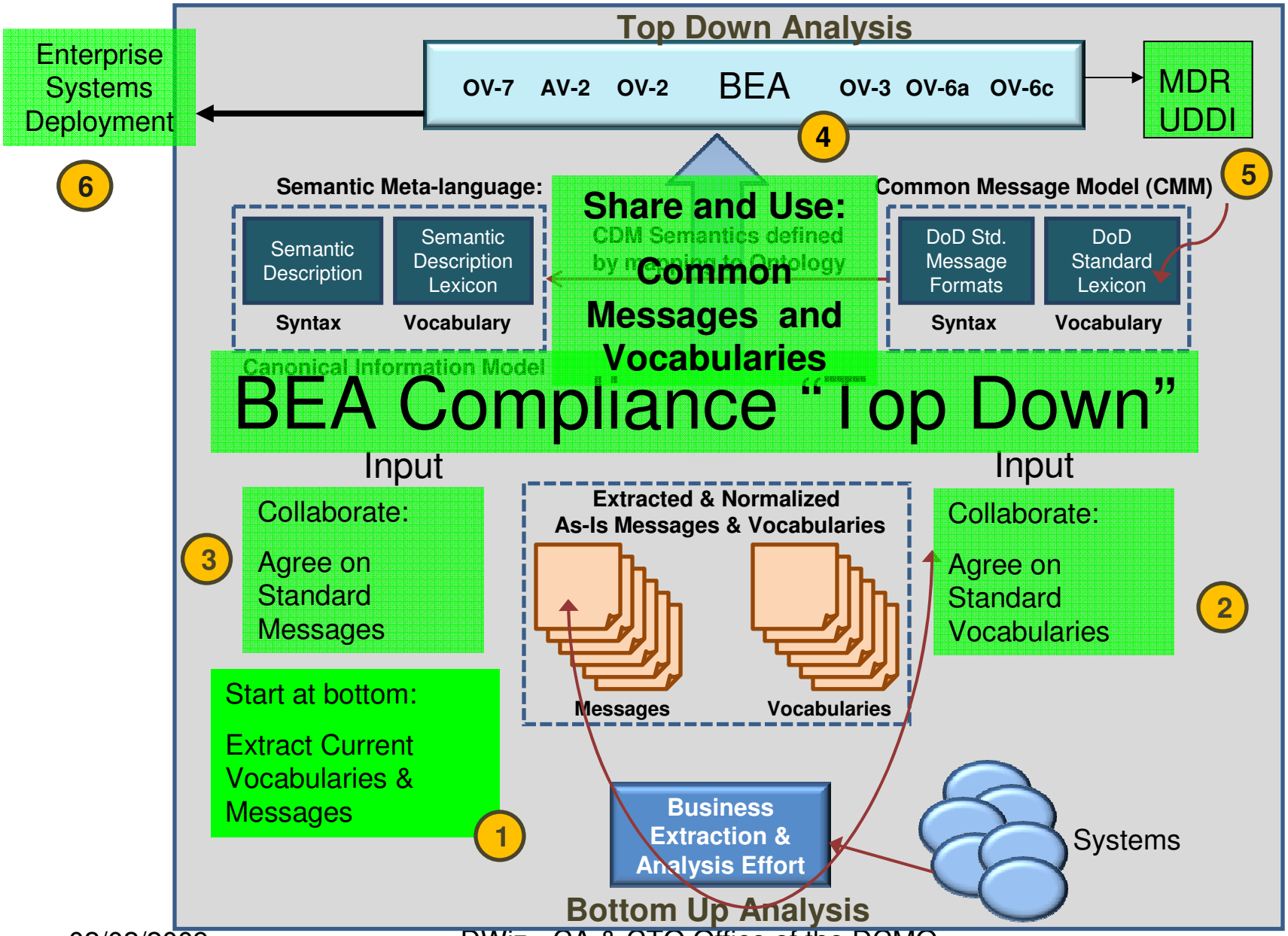
Office of the CMO

Chief Architect

Chief Technical Officer

February 10, 2009

Interoperability via Common Messages and Vocabulary



The Next Step: COIs

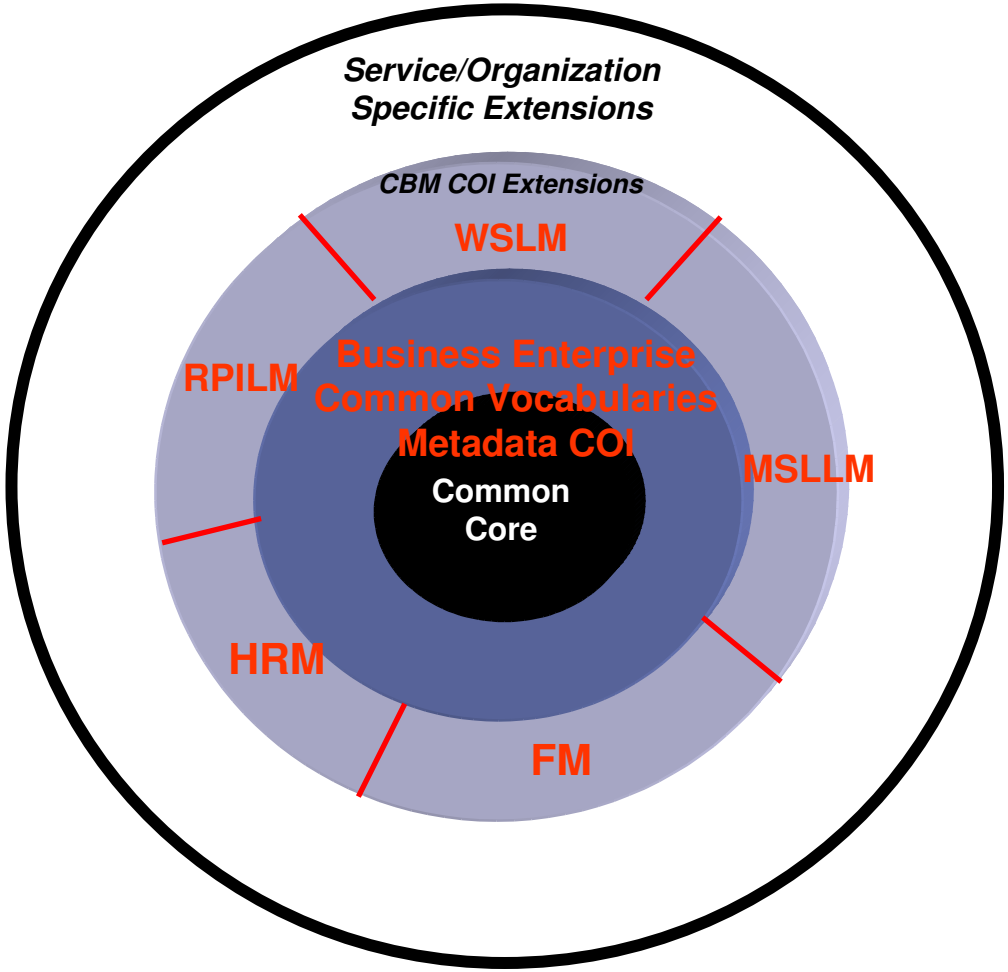
"COIs … come together to **address a specific information sharing mission or challenge** that the COI can solve by exposing and sharing data."—DoD 8320.2G, April 12, 2006

"A collaborative group of users that must exchange information in pursuit of its shared goals, interests, missions, or business processes and therefore must have shared vocabulary for the information exchanges."—*DoD 8320.2, December 2, 2004*

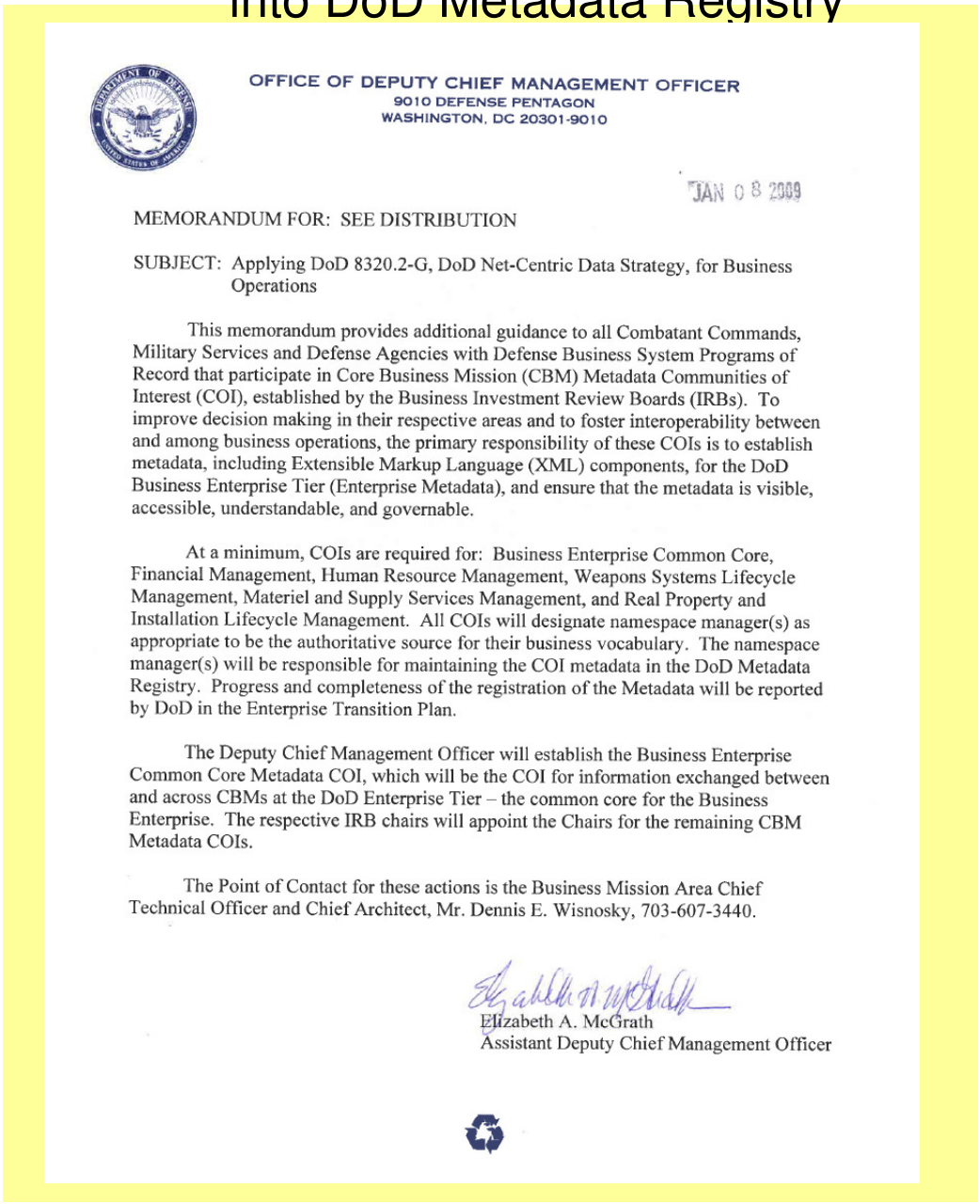
From NII / COI Training Sept 2007

Common Core Data Schema

- **Applies policy & guidance from DoD CIO & NII**
- **Common Core data schema to enable information sharing for Business Operations**
  - **Describes Common Vocabulary for Business Operations**
  - **Business Common Vocabulary chapters for each CBM**
  - **May be extended by CBM down**



Policy for registration of Business Enterprise Metadata into DoD Metadata Registry



OFFICE OF DEPUTY CHIEF MANAGEMENT OFFICER
9010 DEFENSE PENTAGON
WASHINGTON, DC 20301-9010

JAN 08 2009

MEMORANDUM FOR: SEE DISTRIBUTION

SUBJECT: Applying DoD 8320.2-G, DoD Net-Centric Data Strategy, for Business Operations

This memorandum provides additional guidance to all Combatant Commands, Military Services and Defense Agencies with Defense Business System Programs of Record that participate in Core Business Mission (CBM) Metadata Communities of Interest (COI), established by the Business Investment Review Boards (IRBs). To improve decision making in their respective areas and to foster interoperability between and among business operations, the primary responsibility of these COIs is to establish metadata, including Extensible Markup Language (XML) components, for the DoD Business Enterprise Tier (Enterprise Metadata), and ensure that the metadata is visible, accessible, understandable, and governable.

At a minimum, COIs are required for: Business Enterprise Common Core, Financial Management, Human Resource Management, Weapons Systems Lifecycle Management, Materiel and Supply Services Management, and Real Property and Installation Lifecycle Management. All COIs will designate namespace manager(s) as appropriate to be the authoritative source for their business vocabulary. The namespace manager(s) will be responsible for maintaining the COI metadata in the DoD Metadata Registry. Progress and completeness of the registration of the Metadata will be reported by DoD in the Enterprise Transition Plan.

The Deputy Chief Management Officer will establish the Business Enterprise Common Core Metadata COI, which will be the COI for information exchanged between and across CBMs at the DoD Enterprise Tier – the common core for the Business Enterprise. The respective IRB chairs will appoint the Chairs for the remaining CBM Metadata COIs.

The Point of Contact for these actions is the Business Mission Area Chief Technical Officer and Chief Architect, Mr. Dennis E. Wisnosky, 703-607-3440.

Elizabeth A. McGrath
Assistant Deputy Chief Management Officer

Policy Indicates

- COIs as designated by authority
- COI (through namespace manager) has sole authority to enter data for Business Common Core Metadata (Vocabulary) at Enterprise tier
- Registration Authority for Business Enterprise Common Core Metadata COI for Policy is as delegated by the DCMO
  - **Business Enterprise Common Core Metadata COI** is the Business Operations domain COI for information exchanged between and across CMBA's by Defense Business Systems at the Enterprise Tier. CBMA COIs are aligned to the Defense Business System IRBs
  - The Deputy Undersecretary of Defense, Comptroller
    - **FM CBMA COI**
  - The Deputy Undersecretary of Defense for Personnel and Readiness:
    - **HRM CBMA COI**.
  - The Deputy Undersecretary of Defense for Acquisition, Technology and Logistics:
    - **MSSLM CBMA COI**
    - **RPILM CBMA COI**
    - **WSLM CBMA COI**

Governance Figure with 3 DoD CIO Groups

CIO Guidance
Architecture Standards & Policies
Oversee management of IT portfolios against guidance

Technical
DoD/IC Enterprise Services and Data Engineering Governance

Business Operating Environment
Enterprise Services, BTI, Programs, Defense Business Systems
Data Governance

DBSMC/CMO (DepSecDef)

DoD CIO

CIO/C4 Principals Meeting

DoD CIO Executive Board

DNI CIO Forum

Defense Business Systems Management Committee /DCMO

Military Communications Electronics Board

Business Enterprise Common Core Metadata Community of Interest – Common Core

PSA / Certification Authorities

Investment Review Boards
FM, HRM, RP&ILM, MSLM, WSLM, EIE (EGB)

Business Communities of Interest
Links to Warfighter & Intelligence COIs

- Financial Management
- Human Resource Management
- Real Property & Installations Lifecycle Management
- Material Systems Lifecycle Management
- Weapons Systems Lifecycle Management

Enterprise Guidance Board

Engineering Review Committee (ERG)

Operations Oversight Group

Council of Colonels

EGG Tiger Teams

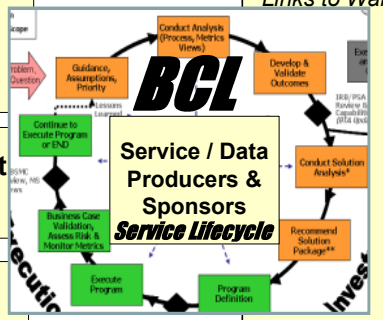
Enterprise Services Acquisition Management Group

Enterprise Services Engineering Review Board

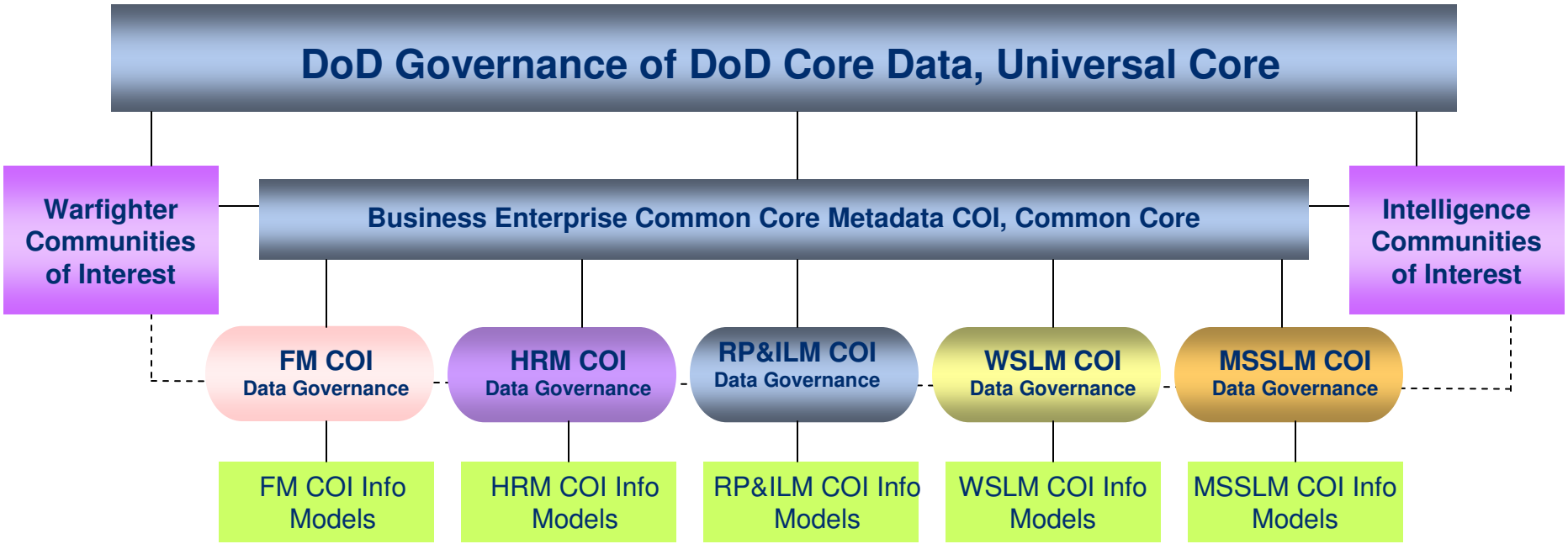
Engineering Working Groups

Enterprise/Component Functional Sponsor

Program-Level Governance
Enterprise or Component including SAEs, PEOs, PMOs and others



Data Governance



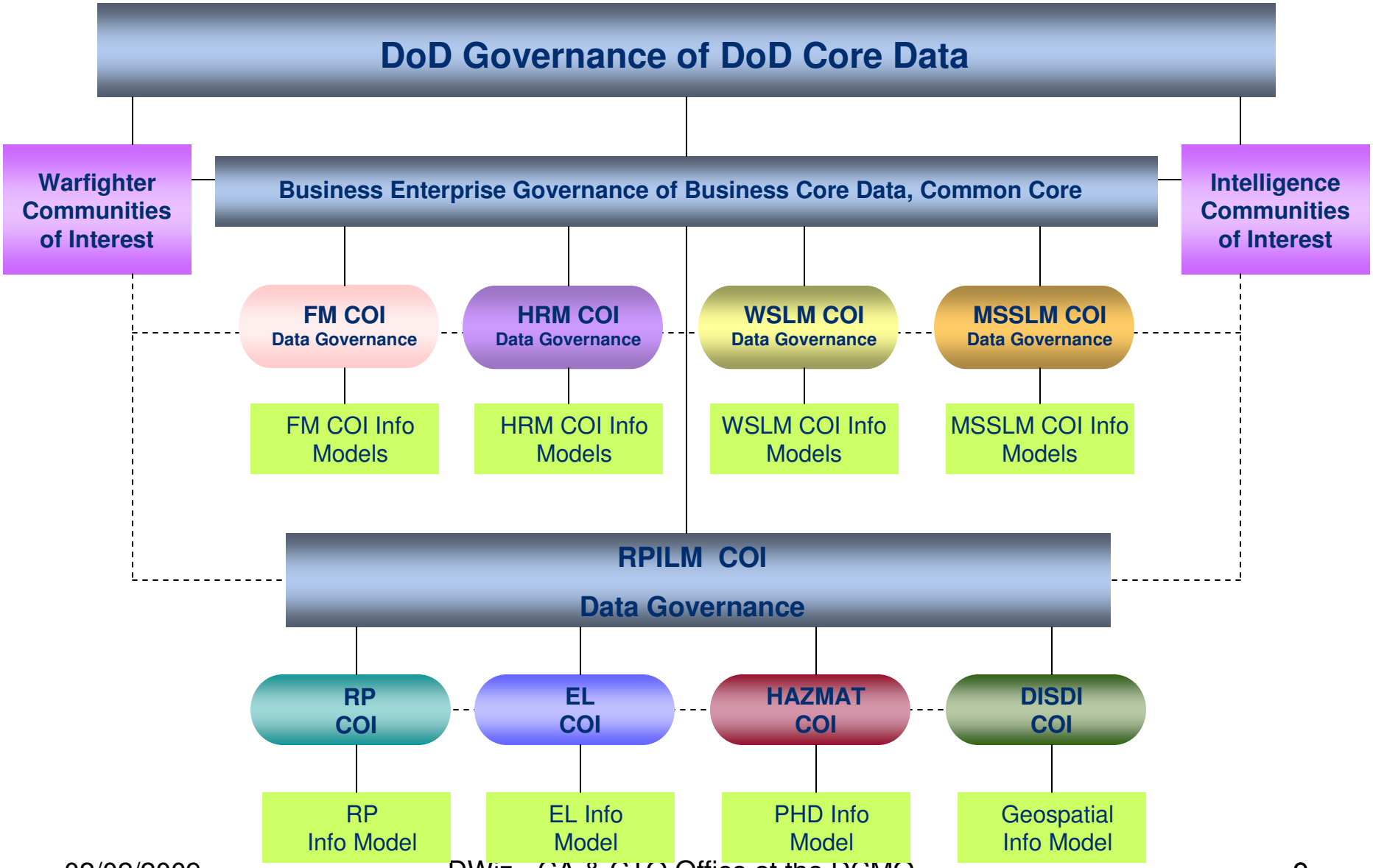
Each COI needs to have "control" over -

- COI Data Assets
- Information Models (Business Language)
- SOA Run Time Products

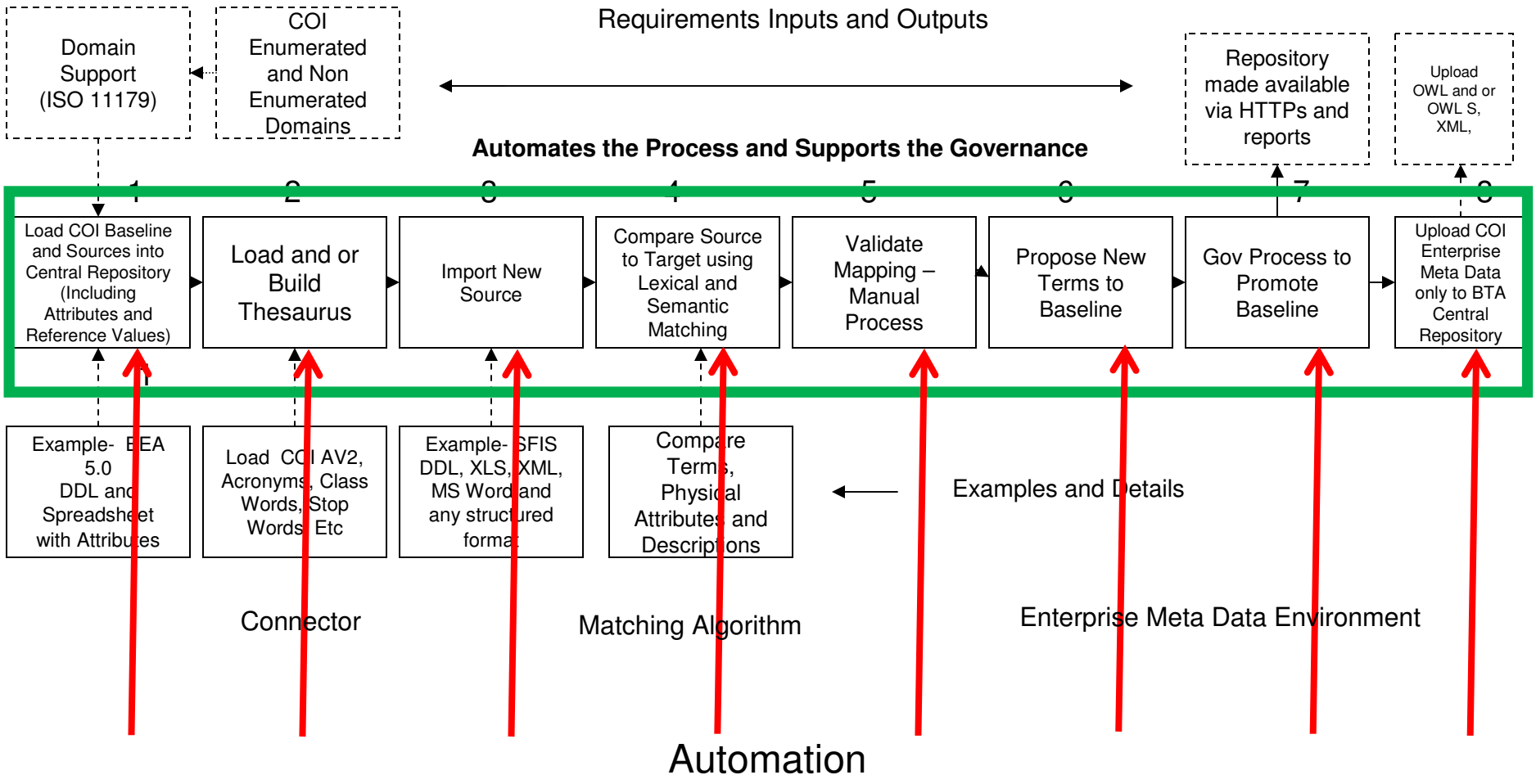
Each CBM needs to have "control" over its own Core Data

Business Operations Enterprise tier needs to have control over Enterprise Core Data

Data Governance – RP&ILM CBMA Example



COI Common Process for Comparing and Building a Common Vocabulary



Matching Approach for COI: # Technical Approach

Register Sources and Transform input records into a common record format

Break terms and descriptions into words and classify them.

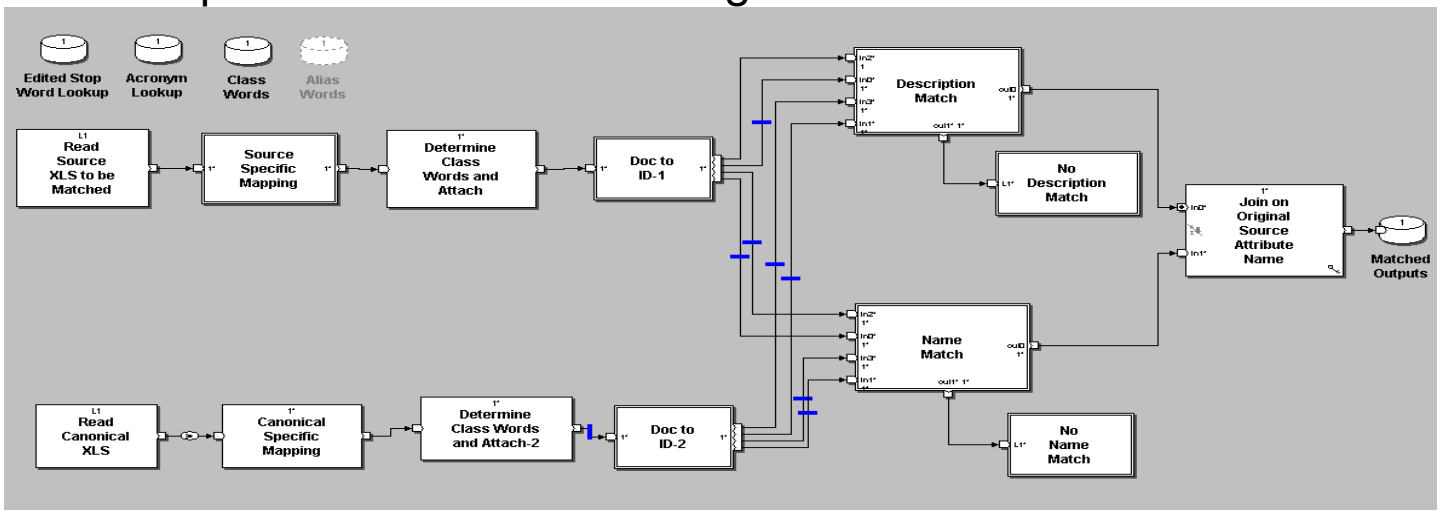
Reduce Term variability:

1. Cleansing
2. Remove Nuisance Words
3. Expand Abbreviations
4. Determine Class Word
5. Expand to Aliases

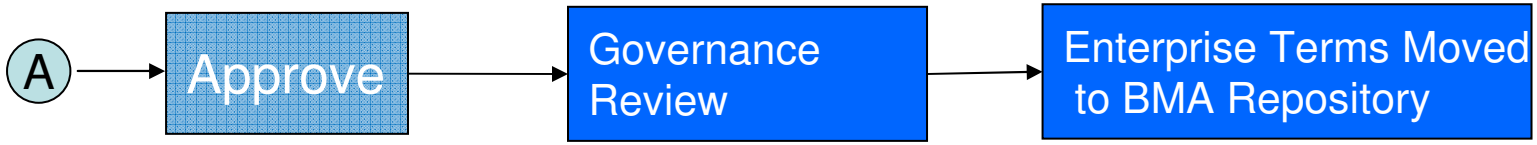
Match Source Terms with Canonical Terms using matching techniques. TF-IDF Fuzzy Linguistic, etc.

User developed rules to grade matches as A, B,C, or F. Use Business Rules Environment to analyze and report.

The Graph that does the Matching .



Matching Approach for COI: Technical Approach (cont.)



User review and approval to accept grades matches and propose new Canonical Terms.

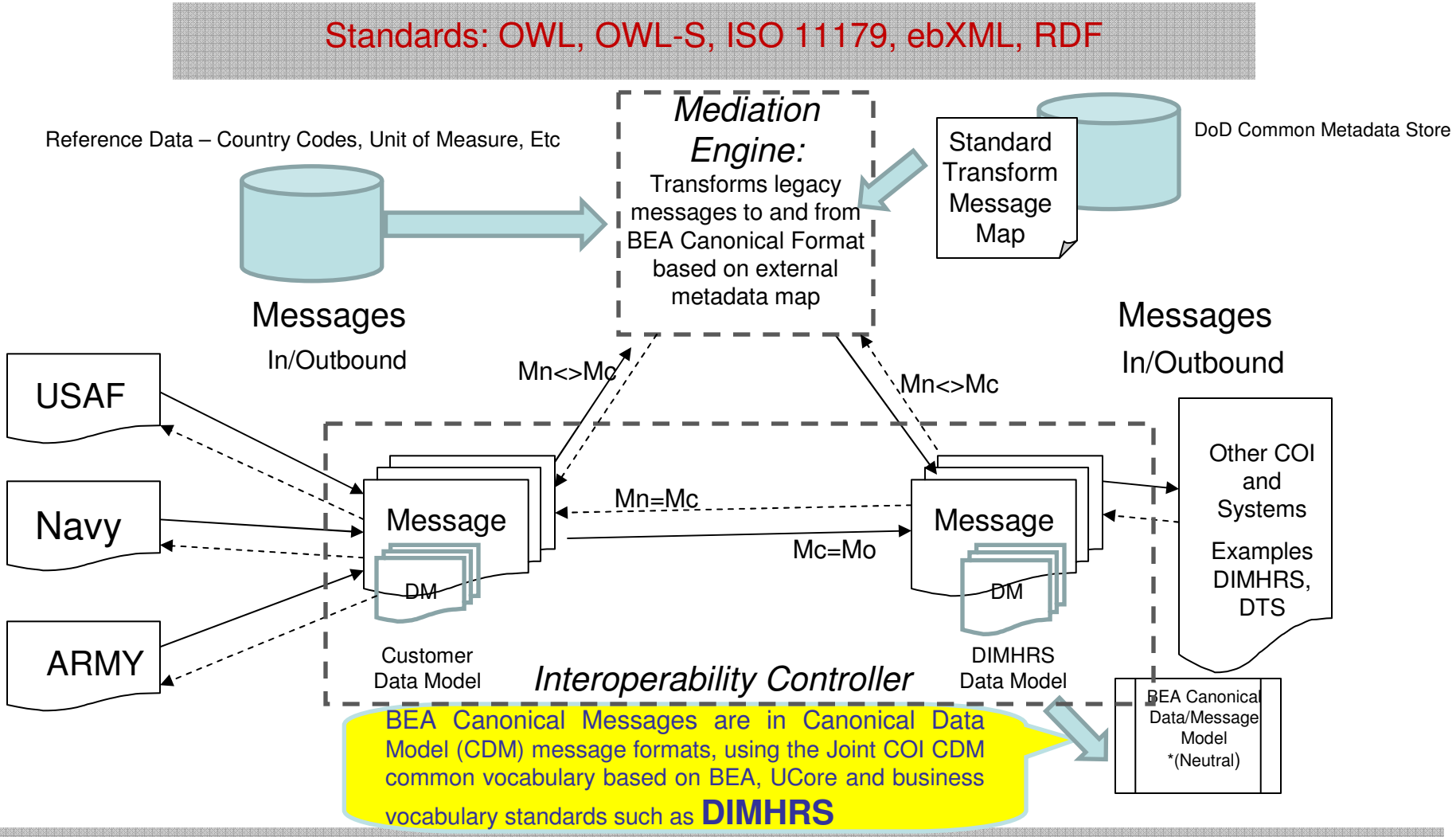
Promoted Terms are reviewed by Governance Board within Meta Data Repository with suggested attributes which are accepted or rejected or sent back for further analysis

Enterprise Terms are promoted as a Enterprise Term within the BMA CV and becomes part of the BEA

Source Attr Name	Source Attr Desc	Source Attr Class Words	Canonical Attr Name	Canonical Attr Desc	Canonical Attr Class Words	Canonical Match Similarity	Match Grade
address type	code that indicates type of address blank or spacepresent address 1 first former address 2second former address 3temporary address		address identifier	the designator that distinguishes one address geopolitical area from another	identifier	0.719838501	C
affiliation code	code that denotes the corporate relationship within a corporate complex structure	code	person ethnic affiliation code	a character string that stands for a recognized cultural group within which the person asserts membership	code	0.657954584	C
affirmative action program contract indicator	the indicator that a business has or has not participated in a previous contract or subcontract subject to the equal opportunity clause of a solicitation	indicator	small business administration criteria affirmative action program indicator	the character string that indicates if the vendor developed or has on file an affirmative action program as required by the secretary of labor values 0 no 1 yes	indicator	0.659983012	C
affirmative action program indicator	the indicator that the business participates in a affirmative action program	indicator	small business administration criteria affirmative action program indicator	the character string that indicates if the vendor developed or has on file an affirmative action program as required by the secretary of labor values 0 no 1 yes	indicator	0.704205623	C
air commodity special handling code	15356 a a code that categorizes a type of shipment unit transported by air	code	shipment air special handling code	a categorization scheme that describes a specific processing requirement of a shipment unit transported by air	code	0.761561492	C
air dimension code	a code that identifies types of dimensional characteristics of air shipments	code	dimension type code	the symbol that stands for the dimension type	code	0.650086279	C

Advanced Mediation Pattern – Runtime how will this be used?

BMA Common Vocabulary and Mediation



Standardized BMA Canonical (Common) Message Model and Mediation driven by Meta Data

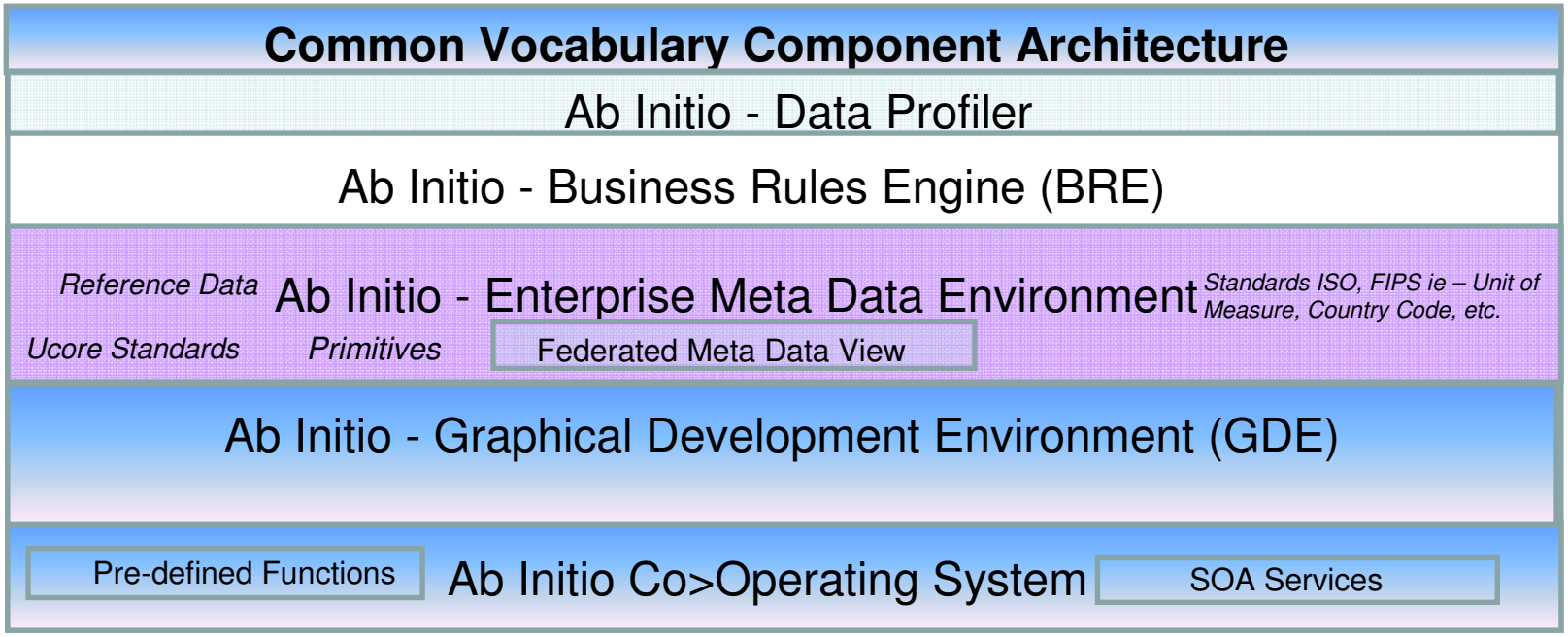
Pipelining, Graphs controlled level paralleling, Data level parallelism is used at all interfaces where possible

Questions?

Backups

Common Vocabulary Component Architecture

Business Transformation Infrastructure (BTI) Common Vocabulary



Matching Approach: Term Frequency–Inverse Document Frequency Method

- The **tf–idf weight** is a statistical measure used to evaluate how important a word is to a document and to a collection of documents. *(Read document as Attribute).*
- The importance of a word increases proportionally to the number of times a word appears in the attribute.
- But it is offset by the frequency of the word in the collection of Attributes in the Canonical Model.
- The TF-IDF weight **diminishes** the weight of words that occur very frequently in the collection and **increases** the weight of words that occur rarely.

Matching Approach: # TF – IDF The Math

The weight vector for document *d* is $\mathbf{v}_d = [w_{1,d}, w_{2,d}, \ldots, w_{N,d}]^T$, where

$$w_{t,d} = \mathrm{tf}_t \cdot \log \frac{|D|}{|\{t \in d\}|}$$

and

- tf_t is term frequency of term *t* in document *d* (a local parameter)
- $\log \frac{|D|}{|\{t \in d\}|}$ is inverse document frequency (a global parameter).

| *D* | is the total number of documents in the document set $|\{t \in d\}|$ is the number of documents containing the term *t*.

Matching Approach: # TF – IDF Example

- An Attribute containing 8 words and the word **branch** appears 2 times. The term frequency (TF) for **branch** is .25 (2/8).
- In the Canonical Model, there are ~3,300 Attributes and assume **branch** appears in 12 of these. Then, the inverse document frequency is calculated as **ln**(3,300 / 12) = 5.61. The TF-IDF score is the product of these quantities: 0.25 * 5.61 = **1.4**.

- An Attribute containing 8 words and the word **code** appears 1 time. The term frequency (TF) for **code** is .125 (1/8).
- In the Canonical Model, there are ~3,300 Attributes and assume **code** appears in 900 of these. Then, the inverse document frequency is calculated as **ln**(3,300 / 900) = 1.99. The TF-IDF score is the product of these quantities: 0.125 * 1.99 = **.16**

- So, the word **branch** with a factor of 1.4 is likely to be more important than the word **code** with a factor of .16.